In many cases, sigmoid function or logistic distribution is widely used in many research when we try to fit model for category dependent variables. But why we use them often? Because there are many alteration like hyper tangent function or restricted linear regression which dependent variable is conditional probability. This can be proved by Markov inequality. Before do it, I'll briefly summarize what is logistic regression and its property.

Definition)Let $Y_{X=x}$ be conditional binary random variable and $p = P(Y_{X=x} = 1)$ be conditional probability. Then the logistic regression is a regression between dependent variable $\log(\frac{p}{1-p})$ and independent variable X as $\log\left(\frac{p}{1-p}\right) = Xb + e \dots (0)$ which b is coefficients and e is error following log normal distribution.

Then $\log\left(\frac{p}{1-p}\right) = Xb + e$ is equivalent to $p = \frac{e^{bx}}{1+e^{bx}}$, which the letter is called logistic distribution. you can see that we need to check that conditional probability is proportional to conditions. If we fit model like (0), then we can determine some threshold x for deciding whether dependent random variable(r.v) is 0 or 1. Of course, this threshold must be defined by cost function. Many people use 0-1 cost function which gives penalty on wrong decision, but this results may be critical when there are some catastrophic rare events. So we need to set cost function before making decision.

Also, we can modify $p = \frac{e^{bx}}{1+e^{bx}}$ by our purpose. For examples, we can add some bias q like $p = \frac{e^{bx}}{1+e^{bx}} + q$ which means there is minimal probability for events no matter what condition X is. Since the probability is proportional to exponent, we need to reduce error term e as much as possible. Or not, probability p has some huge bias and result wrong decision. There are several methods to reduce error. One is that there are no omitted variable in models and another is that applying some techniques for reducing measurement error like instrument variables or else.

We need to check whether there are no omitted variable by domain knowledge. For example, we can check whether 3rd variable or choices have effect on subjects. Suppose you make some recommendation algorithms using logistic regression for Netflix. If the user get some category services provide, then you need to check whether tendency to watch is changeable by another advertisements that is not displayed.

The propostion1 is about markov inequality and I'll use it for some lower bound of probability which is logistic distribution.

proposition1)Let Y be arbitrary non-negative random variable and $a, t > 0$ be real number. Then we know markov inequality below.

$$P(Y > a) \leq \frac{E(Y)}{a}$$

We can substitute $Y = Ratio(Y_{X=x}) = \frac{N(Y_{X=x}=1)}{N(Y_{X=x}=0)}$ and $a = 1 + e^{bx}$ and calculate $P(Y < a)$. Then we

can get logistic distribution as lower bound for probability like (1).

$$P\left(\frac{Ratio(Y_{X=x})}{E(Ratio(Y_{X=x}))} > 1 + \mathrm{e}^{bx}\right) \leq \frac{1}{1+\mathrm{e}^{bx}}$$

$$P\left(\frac{Ratio(Y_{X=x})}{E(Ratio(Y_{X=x}))} \leq 1 + \mathrm{e}^{bx}\right) = P(Ratio(Y_{X=x}) \leq E(Ratio(Y_{X=x}))(1+\mathrm{e}^{bx})) \geq \frac{\mathrm{e}^{bx}}{1+\mathrm{e}^{bx}} \dots (1)$$

(1) means that if we don't know distribution of some r.v $\frac{p(Y_{X=x}=1)}{1-p(Y_{X=x}=1)}$, we can use logistic distribution in worst case. But this requires iterative measurement of binary conditional r.v $Y_{X=x}$ to calculate not only $Ratio(Y_{X=x})$ but also $E(Ratio(Y_{X=x}))$. One situation that satisfies this condition is that you have subjects and measure some variable iteratively for each subjects. For example, YouTube can write not only your decision for whether you click their recommendation videos displayed, but also other's. Then $Ratio(Y_{X=x})$ is calculated for each subjects separately and $E(Ratio(Y_{X=x}))$ is done by all subjects.

You can use model (0) for calculating conditional probabilities but you should consider all hypotheses regression must satisfy. And you can use non-linear regression if you need to. However, you can only use this result for estimating hidden binary variables if you have multiple measurements for defining $E(Ratio(Y_{X=x}))$.